

PATENT COOPERATION TREATY

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

Assistant Commissioner for Patents
United States Patent and Trademark
Office
Box PCT
Washington, D.C. 20231
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 31 March 2000 (31.03.00)	
International application No. PCT/GB99/02517	Applicant's or agent's file reference RSJ05741WO
International filing date (day/month/year) 30 July 1999 (30.07.99)	Priority date (day/month/year) 30 July 1998 (30.07.98)
Applicant STEEL, Samuel, William, Dyne et al	

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:
17 February 2000 (17.02.00)

☐ in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was
☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No.: (41-22) 740.14.35	Authorized officer Juan Cruz Telephone No.: (41-22) 338.83.38
---	---

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

PCT

NOTIFICATION OF TRANSMITTAL OF
THE INTERNATIONAL SEARCH REPORT
OR THE DECLARATION

(PCT Rule 44.1)

To:
GILL JENINGS & EVERY
Broadgate House
7 Eldon Street
LONDON EC2M 7LH
UNITED KINGDOM

DIARIED

RECEIVED

3 MAR 2000

GILL JENNINGS & EVERY

Date of mailing
(day/month/year) 29/02/2000

Applicant's or agent's file reference
RSJ05741W0

FOR FURTHER ACTION See paragraphs 1 and 4 below

International application No.
PCT/GB 99/ 02517

International filing date
(day/month/year) 30/07/1999

Applicant

BRITISH TELECOMMUNICATIONS PLC et al.

1. ☒ The applicant is hereby notified that the International Search Report has been established and is transmitted herewith.

Filing of amendments and statement under Article 19:

The applicant is entitled, if he so wishes, to amend the claims of the International Application (see Rule 46):

When? The time limit for filing such amendments is normally 2 months from the date of transmittal of the International Search Report; however, for more details, see the notes on the accompanying sheet.

Where? Directly to the International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland
Facsimile No.: (41-22) 740.14.35

For more detailed instructions, see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no International Search Report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. ☐ With regard to the protest against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Further action(s):** The applicant is reminded of the following:

Shortly after 18 months from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90bis.1 and 90bis.3, respectively, before the completion of the technical preparations for international publication.

Within 19 months from the priority date, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later).

Within 20 months from the priority date, the applicant must perform the prescribed acts for entry into the national phase before all designated Offices which have not been elected in the demand or in a later election within 19 months from the priority date or could not be elected because they are not bound by Chapter II.

Name and mailing address of the International Searching Authority



European Patent Office, P.B. 5818 Patentlaan 2
NL-2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Lucia Van Pinxteren

NOTES TO FORM PCT/ISA/220

These Notes are intended to give the basic instructions concerning the filing of amendments under article 19. The Notes are based on the requirements of the Patent Cooperation Treaty, the Regulations and the Administrative Instructions under that Treaty. In case of discrepancy between these Notes and those requirements, the latter are applicable. For more detailed information, see also the PCT Applicant's Guide, a publication of WIPO.

In these Notes, "Article", "Rule", and "Section" refer to the provisions of the PCT, the PCT Regulations and the PCT Administrative Instructions, respectively.

INSTRUCTIONS CONCERNING AMENDMENTS UNDER ARTICLE 19

The applicant has, after having received the international search report, one opportunity to amend the claims of the international application. It should however be emphasized that, since all parts of the international application (claims, description and drawings) may be amended during the international preliminary examination procedure, there is usually no need to file amendments of the claims under Article 19 except where, e.g. the applicant wants the latter to be published for the purposes of provisional protection or has another reason for amending the claims before international publication. Furthermore, it should be emphasized that provisional protection is available in some States only.

What parts of the international application may be amended?

Under Article 19, only the claims may be amended.

During the international phase, the claims may also be amended (or further amended) under Article 34 before the International Preliminary Examining Authority. The description and drawings may only be amended under Article 34 before the International Examining Authority.

Upon entry into the national phase, all parts of the international application may be amended under Article 28 or, where applicable, Article 41.

When?

Within 2 months from the date of transmittal of the international search report or 16 months from the priority date, whichever time limit expires later. It should be noted, however, that the amendments will be considered as having been received on time if they are received by the International Bureau after the expiration of the applicable time limit but before the completion of the technical preparations for international publication (Rule 46.1).

Where not to file the amendments?

The amendments may only be filed with the International Bureau and not with the receiving Office or the International Searching Authority (Rule 46.2).

Where a demand for international preliminary examination has been/is filed, see below.

How?

Either by cancelling one or more entire claims, by adding one or more new claims or by amending the text of one or more of the claims as filed.

A replacement sheet must be submitted for each sheet of the claims which, on account of an amendment or amendments, differs from the sheet originally filed.

All the claims appearing on a replacement sheet must be numbered in Arabic numerals. Where a claim is cancelled, no renumbering of the other claims is required. In all cases where claims are renumbered, they must be renumbered consecutively (Administrative Instructions, Section 205(b)).

The amendments must be made in the language in which the international application is to be published.

What documents must/may accompany the amendments?

Letter (Section 205(b)):

The amendments must be submitted with a letter.

The letter will not be published with the international application and the amended claims. It should not be confused with the "Statement under Article 19(1)" (see below, under "Statement under Article 19(1)").

The letter must be in English or French, at the choice of the applicant. However, if the language of the international application is English, the letter must be in English; if the language of the international application is French, the letter must be in French.

NOTES TO FORM PCT/ISA/220 (continued)

The letter must indicate the differences between the claims as filed and the claims as amended. It must, in particular, indicate, in connection with each claim appearing in the international application (it being understood that identical indications concerning several claims may be grouped), whether

- (i) the claim is unchanged;
- (ii) the claim is cancelled;
- (iii) the claim is new;
- (iv) the claim replaces one or more claims as filed;
- (v) the claim is the result of the division of a claim as filed.

The following examples illustrate the manner in which amendments must be explained in the accompanying letter:

1. [Where originally there were 48 claims and after amendment of some claims there are 51]:
"Claims 1 to 29, 31, 32, 34, 35, 37 to 48 replaced by amended claims bearing the same numbers; claims 30, 33 and 36 unchanged; new claims 49 to 51 added."
2. [Where originally there were 15 claims and after amendment of all claims there are 11]:
"Claims 1 to 15 replaced by amended claims 1 to 11."
3. [Where originally there were 14 claims and the amendments consist in cancelling some claims and in adding new claims]:
"Claims 1 to 6 and 14 unchanged; claims 7 to 13 cancelled; new claims 15, 16 and 17 added." or
"Claims 7 to 13 cancelled; new claims 15, 16 and 17 added; all other claims unchanged."
4. [Where various kinds of amendments are made]:
"Claims 1-10 unchanged; claims 11 to 13, 18 and 19 cancelled; claims 14, 15 and 16 replaced by amended claim 14; claim 17 subdivided into amended claims 15, 16 and 17; new claims 20 and 21 added."

"Statement under article 19(1)" (Rule 46.4)

The amendments may be accompanied by a statement explaining the amendments and indicating any impact that such amendments might have on the description and the drawings (which cannot be amended under Article 19(1)).

The statement will be published with the international application and the amended claims.

It must be in the language in which the international application is to be published.

It must be brief, not exceeding 500 words if in English or if translated into English.

It should not be confused with and does not replace the letter indicating the differences between the claims as filed and as amended. It must be filed on a separate sheet and must be identified as such by a heading, preferably by using the words "Statement under Article 19(1)."

It may not contain any disparaging comments on the international search report or the relevance of citations contained in that report. Reference to citations, relevant to a given claim, contained in the international search report may be made only in connection with an amendment of that claim.

Consequence if a demand for international preliminary examination has already been filed

If, at the time of filing any amendments and any accompanying statement, under Article 19, a demand for international preliminary examination has already been submitted, the applicant must preferably, at the time of filing the amendments (and any statement) with the International Bureau, also file with the International Preliminary Examining Authority a copy of such amendments (and of any statement) and, where required, a translation of such amendments for the procedure before that Authority (see Rules 55.3(a) and 62.2, first sentence). For further information, see the Notes to the demand form (PCT/IPEA/401).

Consequence with regard to translation of the international application for entry into the national phase

The applicant's attention is drawn to the fact that, upon entry into the national phase, a translation of the claims as amended under Article 19 may have to be furnished to the designated/elected Offices, instead of, or in addition to, the translation of the claims as filed.

For further details on the requirements of each designated/elected Office, see Volume II of the PCT Applicant's Guide.

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference RSJ05741W0	FOR FURTHER ACTION see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/GB 99/ 02517	International filing date (day/month/year) 30/07/1999	(Earliest) Priority Date (day/month/year) 30/07/1998
Applicant BRITISH TELECOMMUNICATIONS PLC et al.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 6 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

- a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1 (b)).

- b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

☐ contained in the international application in written form.

☐ filed together with the international application in computer readable form.

☐ furnished subsequently to this Authority in written form.

☐ furnished subsequently to this Authority in computer readable form.

☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. ☐ Certain claims were found unsearchable (See Box I).

3. ☒ Unity of invention is lacking (see Box II).

4. With regard to the title,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No.

☒ as suggested by the applicant.

☐ because the applicant failed to suggest a figure.

☐ because this figure better characterizes the invention.

1
☐ None of the figures.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/02517

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LUNIEWSKI, A. ET AL: "RUFUS: Managing Semi-structured information" ALMADEN COMPUTER SCIENCE SHOW AND TELL, ALMADEN RESEARCH CENTER, SAN JOSE, CALIFORNIA, USA, 4 October 1994 (1994-10-04), XP002088419 http://www.almaden.ibm.com/cs/showtell/rufus/overview.html the whole document --- -/--	1-9, 12-20

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents :

A document defining the general state of the art which is not considered to be of particular relevance

E earlier document but published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

8 February 2000

Date of mailing of the international search report

29.02.00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Abbing, R

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HAMMER J ET AL: "Extracting semistructured information from the Web" PROCEEDINGS OF THE WORKSHOP ON MANAGEMENT OF SEMI-STRUCTURED DATA, PROCEEDINGS OF WORKSHOP ON MANAGEMENT OF SEMI-STRUCTURED DATA, TUCSON, AZ, USA, 16 MAY 1997, pages 18-25, XP002099172 1997, Murray Hill, NJ, USA, AT & T Labs - Research, USA page 23, line 29 -page 23, line 44 ---	11,12, 21,22
X	EP 0 522 591 A (MITSUBISHI ELECTRIC CORP) 13 January 1993 (1993-01-13) abstract column 1, line 22 -column 5, line 41 claims ---	10,11, 21,22
X	FRANK SHOU-CHENG TSENG ET AL: "EXTENDING THE E-R CONCEPTS TO CAPTURE NATURAL LANGUAGE SEMANTICS FOR DATABASE ACCESS" PROCEEDINGS OF THE INTERNATIONAL COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE. (COMPSAC),US,LOS ALAMITOS, IEEE. COMP. SOC. PRESS, vol. CONF. 15, 1991, pages 30-35, XP000260517 the whole document ---	23,34
Y	HUFFMAN S ET AL: "Notes Explorer: entity-based retrieval in shared, semi-structured information spaces" PROCEEDINGS OF THE 1996 ACM CIKM. INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, PROCEEDINGS OF 5TH INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, ROCKVILLE, MD, USA, 12-16 NOV. 1996, pages 99-106, XP002088420 ISBN 0-89791-873-8, 1996, New York, NY, USA, ACM, USA the whole document ---	1-9, 12-20
Y	KOPEC G E ET AL: "Document image decoding using Markov source models" CHARACTER RECOGNITION TECHNOLOGIES, SAN JOSE, CA, USA, 1-2 FEB. 1993, vol. 1906, pages 134-145, XP002088421 ISSN 0277-786X, Proceedings of the SPIE - The International Society for Optical Engineering, 1993, USA page 140, line 1 -page 144, line 10 ---	1-9, 12-20
Y	US 5 377 103 A (LAMBERTI DONNA M ET AL) 27 December 1994 (1994-12-27) the whole document ---	23-37
	---	-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 671 425 A (SUEMATSU HIROSHI) 23 September 1997 (1997-09-23) abstract column 4, line 55 -column 6, line 27 figures 1,2A,2B ---	23-37
A	AIRI SALMINEN ET AL: "FROM TEXT TO HYPERTEXT BY INDEXING" ACM TRANSACTIONS ON INFORMATION SYSTEMS, vol. 13, no. 1, January 1995 (1995-01), pages 69-99, XP000501182 abstract page 82, line 16 -page 90, line 7 figures 7-11 ---	1-9, 12-20
A	MCHUGH J ET AL: "LORE: A DATABASE MANAGEMENT SYSTEM FOR SEMISTRUCTURED DATA" SIGMOD RECORD,US,ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, vol. 26, no. 3, 1 September 1997 (1997-09-01), pages 54-66, XP000701384 the whole document ---	10,11, 21,22
A	EP 0 268 367 A (NORTHERN TELECOM LTD) 25 May 1988 (1988-05-25) page 3, line 5 -page 3, line 25 page 3, line 41 -page 4, line 4 -----	23,34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 99/02517

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0522591	A	13-01-1993	JP 2596260 B	02-04-1997
			JP 4348470 A	03-12-1992
			US 5442780 A	15-08-1995

US 5377103	A	27-12-1994	NONE	

US 5671425	A	23-09-1997	US 5418716 A	23-05-1995

EP 0268367	A	25-05-1988	CA 1265871 A	13-02-1990
			JP 63155325 A	28-06-1988

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 99/02517

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-9, 12-20

Method and an apparatus to generate an index to a database containing information in a semi-structured format.

2. Claims: 10,11,21,22

Method and apparatus to identify items in an indexed database consisting of but not restricted to information in a semi-structured format.

3. Claims: 23-37

A method to access a database by processing an input request.



P. B. 5818 - Patentlaan 2
2280 HV Rijswijk (ZH)
☎ (070) 3 40 20 40
TX 31651 epo nl
FAX (070) 3 40 30 16

Europäisches
Patentamt

Zweigstelle
in Den Haag
Recherchen-
abteilung

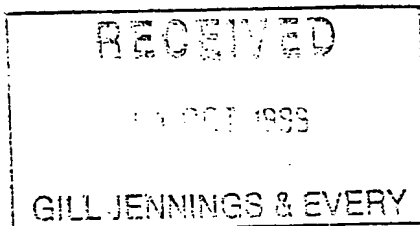
European
Patent Office

Branch at
The Hague
Search
division

Office européen
des brevets

Département à
La Haye
Division de la
recherche

Skone James, Robert Edmund
GILL JENNINGS & EVERY
Broadgate House
7 Eldon Street
London EC2M 7LH
GRANDE BRETAGNE



DIARIED

Datum/Date

08. 10. 1999

Zeichen/Ref./Réf. RSJ05741EP	Anmeldung Nr./Application No./Demande n°/Patent Nr./Patent No./Brevet n°. 98306106.0 ✓
Anmelder/Applicant/Demandeur/Patentinhaber/Proprietor/Titulaire BRITISH TELECOMMUNICATIONS public limited company	

COMMUNICATION

The European Patent Office herewith transmits

- ☒ the European search report
- ☐ the declaration under Rule 45 EPC
- ☐ the partial European search report under Rule 45 EPC
- ☐ the supplementary European search report concerning the international application under Article 157(2) EPC relating to the above-mentioned European patent application. Copies of the documents cited in the search report are enclosed.

The following specifications given by the applicant have been approved by the Search Division :

- ☒ Abstract
- ☒ Title
- ☒ Figure
- ☐ The abstract was modified by the Search Division and the definitive text is attached to this communication.
- ☐ The following figure will be published with the abstract, since the Search Division considers that it better characterises the invention than the one indicated by the applicant.

Figure:

- ☒ Additional copy(copies) of the documents cited in the European search report.

REFUND OF THE SEARCH FEE

If applicable under Article 10 Rules relating to fees, a separate communication from the Receiving Section on the refund of the search fee will be sent later.



EPO Form 1507 02.93



DOCUMENTS CONSIDERED TO BE RELEVANT															
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION												
X	LUNIEWSKI, A. ET AL: "RUFUS: Managing Semi-structured information" ALMADEN COMPUTER SCIENCE SHOW AND TELL, ALMADEN RESEARCH CENTER, SAN JOSE, CALIFORNIA, USA, 4 October 1994 (1994-10-04), XP002088419 http://www.almaden.ibm.com/cs/showtell/rufus/overview.html * the whole document * ---	1-9, 12-20	G06F17/30												
X	HAMMER J ET AL: "Extracting semistructured information from the Web" PROCEEDINGS OF THE WORKSHOP ON MANAGEMENT OF SEMI-STRUCTURED DATA, PROCEEDINGS OF WORKSHOP ON MANAGEMENT OF SEMI-STRUCTURED DATA, TUCSON, AZ, USA, 16 MAY 1997, pages 18-25, XP002099172 1997, Murray Hill, NJ, USA, AT & T Labs - Research, USA * page 23, line 29 - page 23, line 44 * ---	10,11, 21,22													
X	EP 0 522 591 A (MITSUBISHI ELECTRIC CORP) 13 January 1993 (1993-01-13) * abstract * * column 1, line 22 - column 5, line 41 * * claims * --- -/--	10,11, 21,22	TECHNICAL FIELDS SEARCHED G06F												
The present search report has been drawn up for all claims															
Place of search THE HAGUE		Date of completion of the search 7 July 1999	Examiner Abbing, R												
<table border="0"><tr><td>CATEGORY OF CITED DOCUMENTS</td><td>T : theory or principle underlying the invention</td></tr><tr><td>X : particularly relevant if taken alone</td><td>E : earlier patent document, but published on, or after the filing date</td></tr><tr><td>Y : particularly relevant if combined with another document of the same category</td><td>D : document cited in the application</td></tr><tr><td>A : technological background</td><td>L : document cited for other reasons</td></tr><tr><td>O : non-written disclosure</td><td>& : member of the same patent family, corresponding document</td></tr><tr><td>P : intermediate document</td><td></td></tr></table>				CATEGORY OF CITED DOCUMENTS	T : theory or principle underlying the invention	X : particularly relevant if taken alone	E : earlier patent document, but published on, or after the filing date	Y : particularly relevant if combined with another document of the same category	D : document cited in the application	A : technological background	L : document cited for other reasons	O : non-written disclosure	& : member of the same patent family, corresponding document	P : intermediate document	
CATEGORY OF CITED DOCUMENTS	T : theory or principle underlying the invention														
X : particularly relevant if taken alone	E : earlier patent document, but published on, or after the filing date														
Y : particularly relevant if combined with another document of the same category	D : document cited in the application														
A : technological background	L : document cited for other reasons														
O : non-written disclosure	& : member of the same patent family, corresponding document														
P : intermediate document															



DOCUMENTS CONSIDERED TO BE RELEVANT															
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION												
Y	HUFFMAN S ET AL: "Notes Explorer: entity-based retrieval in shared, semi-structured information spaces" PROCEEDINGS OF THE 1996 ACM CIKM. INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, PROCEEDINGS OF 5TH INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, ROCKVILLE, MD, USA, 12-16 NOV. 1996, pages 99-106, XP002088420 ISBN 0-89791-873-8, 1996, New York, NY, USA, ACM, USA * the whole document *	1-9, 12-20													
Y	KOPEC G E ET AL: "Document image decoding using Markov source models" CHARACTER RECOGNITION TECHNOLOGIES, SAN JOSE, CA, USA, 1-2 FEB. 1993, vol. 1906, pages 134-145, XP002088421 ISSN 0277-786X, Proceedings of the SPIE - The International Society for Optical Engineering, 1993, USA * page 140, line 1 - page 144, line 10 *	1-9, 12-20	TECHNICAL FIELDS SEARCHED												
A	AIRI SALMINEN ET AL: "FROM TEXT TO HYPERTEXT BY INDEXING" ACM TRANSACTIONS ON INFORMATION SYSTEMS, vol. 13, no. 1, January 1995 (1995-01), pages 69-99, XP000501182 * abstract * * page 82, line 16 - page 90, line 7 * * figures 7-11 * --- -/--	1-9, 12-20													
The present search report has been drawn up for all claims															
Place of search THE HAGUE		Date of completion of the search 7 July 1999	Examiner Abbing, R												
<table border="0"><tr><td>CATEGORY OF CITED DOCUMENTS</td><td>T : theory or principle underlying the invention</td></tr><tr><td>X : particularly relevant if taken alone</td><td>E : earlier patent document, but published on, or after the filing date</td></tr><tr><td>Y : particularly relevant if combined with another document of the same category</td><td>D : document cited in the application</td></tr><tr><td>A : technological background</td><td>L : document cited for other reasons</td></tr><tr><td>O : non-written disclosure</td><td>& : member of the same patent family, corresponding document</td></tr><tr><td>P : intermediate document</td><td></td></tr></table>				CATEGORY OF CITED DOCUMENTS	T : theory or principle underlying the invention	X : particularly relevant if taken alone	E : earlier patent document, but published on, or after the filing date	Y : particularly relevant if combined with another document of the same category	D : document cited in the application	A : technological background	L : document cited for other reasons	O : non-written disclosure	& : member of the same patent family, corresponding document	P : intermediate document	
CATEGORY OF CITED DOCUMENTS	T : theory or principle underlying the invention														
X : particularly relevant if taken alone	E : earlier patent document, but published on, or after the filing date														
Y : particularly relevant if combined with another document of the same category	D : document cited in the application														
A : technological background	L : document cited for other reasons														
O : non-written disclosure	& : member of the same patent family, corresponding document														
P : intermediate document															

3

EPO FORM 1503 03/82 (P04/C01)



DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION
A	MCHUGH J ET AL: "LORE: A DATABASE MANAGEMENT SYSTEM FOR SEMISTRUCTURED DATA" SIGMOD RECORD, vol. 26, no. 3, 1 September 1997 (1997-09-01), pages 54-66, XP000701384 * the whole document * -----	10,11, 21,22	
			TECHNICAL FIELDS SEARCHED
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 7 July 1999	Examiner Abbing, R
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

3

EPO FORM 1503 03.02 (P04C01)



CLAIMS INCURRING FEES

The present European patent application comprised at the time of filing more than ten claims.

- ☐ Only part of the claims have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims and for those claims for which claims fees have been paid, namely claim(s):
- ☐ No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims.

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

see sheet B

- ☒ All further search fees have been paid within the fixed time limit. The present European search report has been drawn up for all claims.
- ☐ As all searchable claims could be searched without effort justifying an additional fee, the Search Division did not invite payment of any additional fee.
- ☐ Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respect of which search fees have been paid, namely claims:
- ☐ None of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims, namely claims:



The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

1. Claims: 1-9, 12-20

Method and an apparatus to generate an index to a database containing information in a semi-structured format.

2. Claims: 10,11,21,22

Method and apparatus to identify items in an indexed database consisting of but not restricted to information in a semi-structured format.

EP 98 30 6106

07-07-1999

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0522591 A	13-01-1993	JP 2596260 B	02-04-1997
		JP 4348470 A	03-12-1992
		US 5442780 A	15-08-1995

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference RSJ05741WO		FOR FURTHER ACTION see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/GB 99/ 02517	International filing date (day/month/year) 30/07/1999	(Earliest) Priority Date (day/month/year) 30/07/1998	
Applicant BRITISH TELECOMMUNICATIONS PLC et al.			

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 6 sheets.



It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

- a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.



the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

- b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :



contained in the international application in written form.



filed together with the international application in computer readable form.



furnished subsequently to this Authority in written form.



furnished subsequently to this Authority in computer readable form.



the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.



the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2.



Certain claims were found unsearchable (See Box I).

3.



Unity of invention is lacking (see Box II).

4. With regard to the title,



the text is approved as submitted by the applicant.



the text has been established by this Authority to read as follows:

5. With regard to the abstract,



the text is approved as submitted by the applicant.



the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No.

as suggested by the applicant.



because the applicant failed to suggest a figure.



because this figure better characterizes the invention.

1



None of the figures.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/02517

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LUNIEWSKI, A. ET AL: "RUFUS: Managing Semi-structured information" ALMADEN COMPUTER SCIENCE SHOW AND TELL, ALMADEN RESEARCH CENTER, SAN JOSE, CALIFORNIA, USA, 4 October 1994 (1994-10-04), XP002088419 http://www.almaden.ibm.com/cs/showtell/rufus/overview.html the whole document --- -/--	1-9, 12-20

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

8 February 2000

Date of mailing of the international search report

19.02.00

29.02.00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Abbing, R

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/02517

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HAMMER J ET AL: "Extracting semistructured information from the Web" PROCEEDINGS OF THE WORKSHOP ON MANAGEMENT OF SEMI-STRUCTURED DATA, PROCEEDINGS OF WORKSHOP ON MANAGEMENT OF SEMI-STRUCTURED DATA, TUCSON, AZ, USA, 16 MAY 1997, pages 18-25, XP002099172 1997, Murray Hill, NJ, USA, AT & T Labs - Research, USA page 23, line 29 -page 23, line 44 ---	11,12, 21,22
X	EP 0 522 591 A (MITSUBISHI ELECTRIC CORP) 13 January 1993 (1993-01-13) abstract column 1, line 22 -column 5, line 41 claims ---	10,11, 21,22
X	FRANK SHOU-CHENG TSENG ET AL: "EXTENDING THE E-R CONCEPTS TO CAPTURE NATURAL LANGUAGE SEMANTICS FOR DATABASE ACCESS" PROCEEDINGS OF THE INTERNATIONAL COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE. (COMPSAC),US,LOS ALAMITOS, IEEE. COMP. SOC. PRESS, vol. CONF. 15, 1991, pages 30-35, XP000260517 the whole document ---	23,34
Y	HUFFMAN S ET AL: "Notes Explorer: entity-based retrieval in shared, semi-structured information spaces" PROCEEDINGS OF THE 1996 ACM CIKM. INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, PROCEEDINGS OF 5TH INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, ROCKVILLE, MD, USA, 12-16 NOV. 1996, pages 99-106, XP002088420 ISBN 0-89791-873-8, 1996, New York, NY, USA, ACM, USA the whole document ---	1-9, 12-20
Y	KOPEC G E ET AL: "Document image decoding using Markov source models" CHARACTER RECOGNITION TECHNOLOGIES, SAN JOSE, CA, USA, 1-2 FEB. 1993, vol. 1906, pages 134-145, XP002088421 ISSN 0277-786X, Proceedings of the SPIE - The International Society for Optical Engineering, 1993, USA page 140, line 1 -page 144, line 10 ---	1-9, 12-20
Y	US 5 377 103 A (LAMBERTI DONNA M ET AL) 27 December 1994 (1994-12-27) the whole document ---	23-37

	-/--	

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 99/02517

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 671 425 A (SUEMATSU HIROSHI) 23 September 1997 (1997-09-23) abstract column 4, line 55 -column 6, line 27 figures 1,2A,2B ---	23-37
A	AIRI SALMINEN ET AL: "FROM TEXT TO HYPERTEXT BY INDEXING" ACM TRANSACTIONS ON INFORMATION SYSTEMS, vol. 13, no. 1, January 1995 (1995-01), pages 69-99, XP000501182 abstract page 82, line 16 -page 90, line 7 figures 7-11 ---	1-9, 12-20
A	MCHUGH J ET AL: "LORE: A DATABASE MANAGEMENT SYSTEM FOR SEMISTRUCTURED DATA" SIGMOD RECORD,US,ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, vol. 26, no. 3, 1 September 1997 (1997-09-01), pages 54-66, XP000701384 the whole document ---	10,11, 21,22
A	EP 0 268 367 A (NORTHERN TELECOM LTD) 25 May 1988 (1988-05-25) page 3, line 5 -page 3, line 25 page 3, line 41 -page 4, line 4 -----	23,34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 99/02517

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0522591	A	13-01-1993	JP 2596260 B	02-04-1997
			JP 4348470 A	03-12-1992
			US 5442780 A	15-08-1995

US 5377103	A	27-12-1994	NONE	

US 5671425	A	23-09-1997	US 5418716 A	23-05-1995

EP 0268367	A	25-05-1988	CA 1265871 A	13-02-1990
			JP 63155325 A	28-06-1988

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 99/02517

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-9, 12-20

Method and an apparatus to generate an index to a database containing information in a semi-structured format.

2. Claims: 10,11,21,22

Method and apparatus to identify items in an indexed database consisting of but not restricted to information in a semi-structured format.

3. Claims: 23-37

A method to access a database by processing an input request.



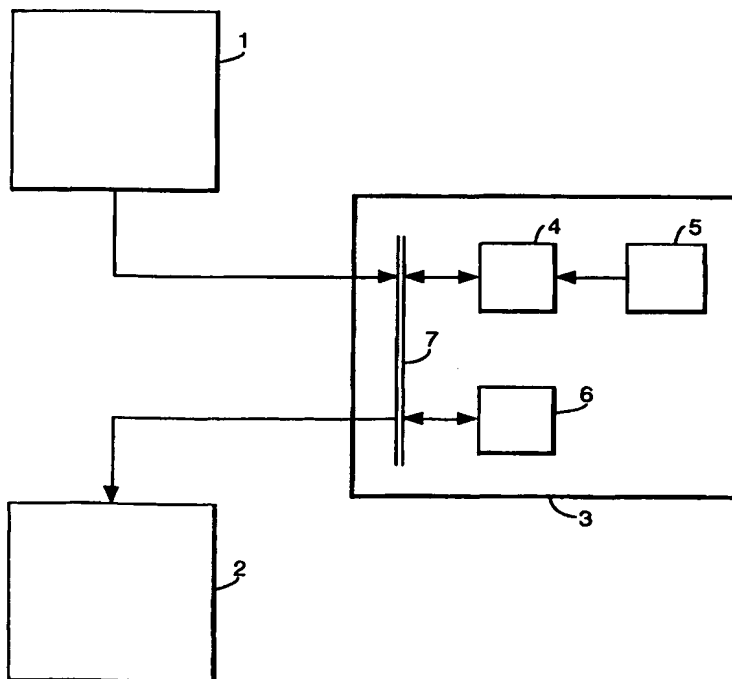
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : G06F 17/30		A3	(11) International Publication Number: WO 00/07117
			(43) International Publication Date: 10 February 2000 (10.02.00)
(21) International Application Number: PCT/GB99/02517		Colchester, Essex CO4 4RR (GB). WOBCKE, Wayne, Raymond [AU/GB]; 12 Sherwood Fields, Kesgrave, Ipswich, Suffolk IP5 2YL (GB). AZVINE, Behnam [GB/GB]; 6 Dodson Vale, Kesgrave, Ipswich, Suffolk IP5 2GT (GB). (74) Agent: GILL JENNINGS & EVERY; Broadgate House, 7 Eldon Street, London EC2M 7LH (GB). (81) Designated States: AU, CA, NZ, SG, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published With international search report. (88) Date of publication of the international search report: 18 May 2000 (18.05.00)	
(22) International Filing Date: 30 July 1999 (30.07.99)			
(30) Priority Data:			
9816648.1 / 30 July 1998 (30.07.98)	GB		
98306106.0 31 July 1998 (31.07.98)	EP		
(71) Applicant (for all designated States except US): BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): STEEL, Samuel, William, Dyne [GB/GB]; The Latch, Church Street, Boxted, Colchester, Essex CO4 3SQ (GB). KRUSCHWITZ, Udo [DE/GB]; 47 Manor Way, Wivenhoe, Colchester, Essex CO7 9LN (GB). WEBB, Nicholas, John [GB/GB]; 3.6A Avon Way House, Avon Way, Colchester, Essex CO4 3TZ (GB). DE ROECK, Anne, Nellie [BE/GB]; 29 School Road, Copford, Colchester, Essex CO6 1BU (GB). SCOTT, Paul, David [GB/GB]; 10 The Hemlocks, Haslingfield, Cambridge CB3 7LE (GB). TURNER, Raymond [GB/GB]; 10 Park Road, Wivenhoe, Colchester, Essex (GB). TSUI, Kwok, Ching [GB/GB]; 138 Berkley Close, Highwoods,			

(54) Title: AN INDEX TO A SEMI-STRUCTURED DATABASE

(57) Abstract

The present invention relates to a method of generating an index (2) to a semi-structured database (1). Semi-structured databases contain a number of items, each of which is stored as a set of semi-structured data including a number of related entries. The presence of these entries are determined by comparing the sets of data to a number of selection criteria, defining one or more predetermined characteristics of various entries. A set of indices is then generated representing a concordance between the determined entries and the respective items.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakistan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/02517

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LUNIEWSKI, A. ET AL: "RUFUS: Managing Semi-structured information" ALMADEN COMPUTER SCIENCE SHOW AND TELL, ALMADEN RESEARCH CENTER, SAN JOSE, CALIFORNIA, USA, 4 October 1994 (1994-10-04), XP002088419 http://www.almaden.ibm.com/cs/showtell/rufus/overview.html the whole document --- -/--	1-9, 12-20

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

8 February 2000

Date of mailing of the international search report

29.02.00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Abbing, R

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/02517

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HAMMER J ET AL: "Extracting semistructured information from the Web" PROCEEDINGS OF THE WORKSHOP ON MANAGEMENT OF SEMI-STRUCTURED DATA, PROCEEDINGS OF WORKSHOP ON MANAGEMENT OF SEMI-STRUCTURED DATA, TUCSON, AZ, USA, 16 MAY 1997, pages 18-25, XP002099172 1997, Murray Hill, NJ, USA, AT & T Labs - Research, USA page 23, line 29 -page 23, line 44 ---	11,12, 21,22
X	EP 0 522 591 A (MITSUBISHI ELECTRIC CORP) 13 January 1993 (1993-01-13) abstract column 1, line 22 -column 5, line 41 claims ---	10,11, 21,22
X	FRANK SHOU-CHENG TSENG ET AL: "EXTENDING THE E-R CONCEPTS TO CAPTURE NATURAL LANGUAGE SEMANTICS FOR DATABASE ACCESS" PROCEEDINGS OF THE INTERNATIONAL COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE. (COMPSAC),US,LOS ALAMITOS, IEEE. COMP. SOC. PRESS, vol. CONF. 15, 1991, pages 30-35, XP000260517 the whole document ---	23,34
Y	HUFFMAN S ET AL: "Notes Explorer: entity-based retrieval in shared, semi-structured information spaces" PROCEEDINGS OF THE 1996 ACM CIKM. INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, PROCEEDINGS OF 5TH INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, ROCKVILLE, MD, USA, 12-16 NOV. 1996, pages 99-106, XP002088420 ISBN 0-89791-873-8, 1996, New York, NY, USA, ACM, USA the whole document ---	1-9, 12-20
Y	KOPEC G E ET AL: "Document image decoding using Markov source models" CHARACTER RECOGNITION TECHNOLOGIES, SAN JOSE, CA, USA, 1-2 FEB. 1993, vol. 1906, pages 134-145, XP002088421 ISSN 0277-786X, Proceedings of the SPIE - The International Society for Optical Engineering, 1993, USA page 140, line 1 -page 144, line 10 ---	1-9, 12-20
Y	US 5 377 103 A (LAMBERTI DONNA M ET AL) 27 December 1994 (1994-12-27) the whole document ---	23-37

	-/--	

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/02517

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 671 425 A (SUEMATSU HIROSHI) 23 September 1997 (1997-09-23) abstract column 4, line 55 -column 6, line 27 figures 1,2A,2B	23-37
A	----- AIRI SALMINEN ET AL: "FROM TEXT TO HYPERTEXT BY INDEXING" ACM TRANSACTIONS ON INFORMATION SYSTEMS, vol. 13, no. 1, January 1995 (1995-01), pages 69-99, XP000501182 abstract page 82, line 16 -page 90, line 7 figures 7-11	1-9, 12-20
A	----- MCHUGH J ET AL: "LORE: A DATABASE MANAGEMENT SYSTEM FOR SEMISTRUCTURED DATA" SIGMOD RECORD,US,ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, vol. 26, no. 3, 1 September 1997 (1997-09-01), pages 54-66, XP000701384 the whole document	10,11, 21,22
A	----- EP 0 268 367 A (NORTHERN TELECOM LTD) 25 May 1988 (1988-05-25) page 3, line 5 -page 3, line 25 page 3, line 41 -page 4, line 4 -----	23,34

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 99/02517

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-9, 12-20

Method and an apparatus to generate an index to a database containing information in a semi-structured format.

2. Claims: 10,11,21,22

Method and apparatus to identify items in an indexed database consisting of but not restricted to information in a semi-structured format.

3. Claims: 23-37

A method to access a database by processing an input request.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 99/02517

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0522591	A	13-01-1993	JP 2596260 B JP 4348470 A US 5442780 A	02-04-1997 03-12-1992 15-08-1995
US 5377103	A	27-12-1994	NONE	
US 5671425	A	23-09-1997	US 5418716 A	23-05-1995
EP 0268367	A	25-05-1988	CA 1265871 A JP 63155325 A	13-02-1990 28-06-1988

REC'D 09 NOV 2000

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference RSJ05741WO	FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/GB99/02517	International filing date (day/month/year) 30/07/1999	Priority date (day/month/year) 30/07/1998
International Patent Classification (IPC) or national classification and IPC G06F17/30		
Applicant BRITISH TELECOMMUNICATIONS PLC et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.


2. This REPORT consists of a total of 12 sheets, including this cover sheet.

- ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of 4 sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☒ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☒ Certain defects in the international application
- VIII ☒ Certain observations on the international application

Date of submission of the demand 17/02/2000	Date of completion of this report 07.11.2000
Name and mailing address of the international preliminary examining authority:  European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Schneider, M Telephone No. +49 89 2399 7509



INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB99/02517

I. Basis of the report

1. This report has been drawn on the basis of (*substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

Description, pages:

1-32 as originally filed

Claims, No.:

1-10 as received on 04/10/2000 with letter of 03/10/2000

Drawings, sheets:

1/3-3/3 as originally filed

2. The amendments have resulted in the cancellation of:

- ☐ the description, pages:
☒ the claims, Nos.: 11-37
☐ the drawings, sheets:

3. ☒ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

see separate sheet

4. Additional observations, if necessary:

IV. Lack of unity of invention

1. In response to the invitation to restrict or pay additional fees the applicant has:

- ☒ restricted the claims.
☒ paid additional fees.
☐ paid additional fees under protest.
☐ neither restricted nor paid additional fees.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB99/02517

2. ☐ This Authority found that the requirement of unity of invention is not complied and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.
3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is
- ☐ complied with.
- ☒ not complied with for the following reasons:
- see separate sheet**
4. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this report:
- ☐ all parts.
- ☒ the parts relating to claims Nos. 1-9, 10.

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims 1-10
	No:	Claims
Inventive step (IS)	Yes:	Claims
	No:	Claims 1-10
Industrial applicability (IA)	Yes:	Claims 1-10
	No:	Claims

2. Citations and explanations

see separate sheet

VII. Certain defects in the international application

The following defects in the form or contents of the international application have been noted:

see separate sheet

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT**

International application No. PCT/GB99/02517

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

see separate sheet

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/GB99/02517

Re Item I

Basis of the report

- (i) Throughout the whole application as originally filed (see e.g. description p. 1, l. 36 - p. 2, l. 5, p. 2, l. 19-27, p. 3, l. 33 - p. 4, l. 32, p. 12, l. 18 - p. 13, l. 5, p. 15, l. 1-4, fig. 2a, originally filed claims 1-3), the impression is conveyed
- that the items handled by the method of the application usually comprise "specific fields" and at least one "free text field", and
 - that entries within specific fields are recognized as such by examining the corresponding field type, and not by searching an entry, corresponding to some predefined format, within the field.

Thus, the feature that all fields of a document are to be treated as given in the characterising part of independent claims 1 and 6 is neither implicitly nor explicitly disclosed in the originally filed application, where this treatment is restricted to free text fields.

Therefore, the application was examined as if claims 1 and 6 would correspond to the original disclosure and were having the following extensions:

Claim 1:

- ..., each record comprising one or more fields having a plurality of characters therein, at least one of the fields being a free text field, the method ...
- ... for each of a plurality of predetermined formats, the analysing step (i) for each free text field further includes ...

Claim 6:

- ..., each record comprising one or more fields having a plurality of characters therein, at least one of the fields being a free text field, the apparatus ...
- ... the processor further includes means for searching each free text field to identify a sequence of characters having a format ...

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/GB99/02517

(ii) The statement of claims 4 and 9, that

"the records are further arranged in groups of records, each group of records being located in a heading field and being identified by at least one heading entry",

was not part of the original disclosure and is not consistent with the description p. 13, l. 28-32. Therefore, the application was examined as if claims 4 and 9 would correspond to the description p. 13, l. 28-32.

(iii) In claim 6 it is stated that "the processor further includes means for searching said field to identify a sequence of characters having a format corresponding to each of a plurality of predetermined formats, ...". This was not part of the original disclosure and is not consistent with the description (see e.g. p. 15, l. 6-24) and claim 1. Therefore, the application was examined as if claim 6 would correspond to the description and claim 1.

Re Item IV

Lack of unity of invention

In agreement with the argumentation and conclusion of the International Search Authority, this International Preliminary Examination Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-9 (based on originally filed claims 1-9, 12-20)

Method and an apparatus to generate an index to a database containing information in a semi- structured format.

2. (originally filed claims: 10,11,21,22)

Methods and apparatus to identify items in an indexed database consisting of but not restricted to information in a semi-structured format.

**INTERNATIONAL PRELIMINARY
EXAMINATION REPORT - SEPARATE SHEET**

International application No. PCT/GB99/02517

3. Claim: 10 (based on originally filed claims 23-37)

A method to access a database by processing an input request.

Re Item V

Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

Reference is made to the following documents:

- D1: HUFFMAN S ET AL: 'Notes Explorer: entity-based retrieval in shared, semi-structured information spaces' PROCEEDINGS OF THE 1996 ACM CIKM. INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, PROCEEDINGS OF 5TH INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, ROCKVILLE, MD, USA, 12-16 NOV. 1996, pages 99-106, XP002088420 ISBN 0-89791-873-8, 1996, New York, NY, USA, ACM, USA
- D2: US-A-5 377 103 (LAMBERTI DONNA M ET AL) 27 December 1994 (1994-12-27)
- D3: FRANK SHOU-CHENG TSENG ET AL: 'EXTENDING THE E-R CONCEPTS TO CAPTURE NATURAL LANGUAGE SEMANTICS FOR DATABASE ACCESS' PROCEEDINGS OF THE INTERNATIONAL COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE. (COMPSAC),US,LOS ALAMITOS, IEEE. COMP. SOC. PRESS, vol. CONF. 15, 1991, pages 30-35, XP000260517

- 1.1 The subject-matter of claim 1, if clarified as set out under Item VIII, does not involve an inventive step in the sense of Article 33(3) PCT, the reasons being as follows:

In D1, p. 99, left col., l. 23-32, p. 100, right col., l. 11-15, p. 105, left col., l. 28-42 and right col., l. 17-21, all features of the preamble of claim 1 are shown within the system "Lotus Notes". In particular, it is clear that a document within a Lotus Notes database comprises short fields and a "blob" field with free-running text, corresponding to a free text field.

In D1, p. 105, left col., l. 28-42, it is proposed to index also entities of "many other categories", as e.g. locations. It also is said that "references to entities can also be found in bodies of free-running text within documents". Explicitly it is suggested

to index entities of the categories "names" and "technical terms" from within the free-running text. These proposals and statements are made in respect to a system called "Notes Explorer", but as Notes Explorer is thought to "allow its users to search for references to entities across a set of [Lotus] Notes databases that are indexed for entity references ahead of time" (see D1, p. 99, right col., l. 34-36) it is implicit that these proposals and statements can, or even should, be applied to Lotus Notes itself.

Evidently entities possibly being of interest for indexing in respective indexes are, besides names of persons, company names and technical term keywords, locations, dates, telephone numbers, email-adresses, bank accounts and dollar amounts, e.g. in order to routinely find all documents related to some postal code or bank accounts (see also D1, p. 100, right col., l. 15-22). In addition, the need for some index, as e.g. the revision or publication date of a document or the email-adress of the author(s), may become evident when the database is already used for some time and the documents stored in it do not have a corresponding short field but include those entities in the free-running text. Then a search for index-related entities in the free-running text seems to be the only way to index these documents correspondingly. The same may happen if a new database is set up for managing legacy documents that in respect to the indexes of the database do not have the adequate set of short fields.

Regarding the way of how to identify entities from within the free-running text, it is well-known to those skilled in the art and also made obvious on D1, p. 100, right col., l. 24-50, that those entities usually correspond to some predefined format and that therefore predefined formats can be used to type recognize and thus to identify the entities.

By using predefined formats to identify the entities within the free-running text in the system of D1 a person skilled in the art arrives at the method of claim 1.

- 1.2 Dependent claims 2 - 5, if clarified as set out under Item VIII, do not appear to contain any additional features which, in combination with the features of any claim to which they refer, meet the requirements of the PCT with respect to inventive step.

- Claim 2:

This claim merely comprises the step of setting an abstract definition.

- Claim 3:

In the system of D1 it is an obvious possibility to identify technical terms within the free running text, which is proposed by D1, p. 105, left col., l. 28-42, by applying the criterion of a word or group of words being part of some list of technical keywords. This may easily be done only after having identified entries corresponding to predefined formats and thus implicitly having defined a free text entry.

- Claim 4:

In a semistructured database in which items can be grouped in clusters according to some characteristics, that are given for each cluster in a cluster description as being valid for all items in the cluster, it is obvious that an index entry that corresponds to one of those characteristics and refers to a cluster description implicitly also indirectly refers to all items in the referred cluster. The generation of explicit index entries to the items is an obvious possibility to avoid the disadvantages of following indirect references during retrieval. The combination of the features of claims 1 and 4 is not inventive as it reveals no unexpected effect.

- Claim 5:

The feature of this claim is obvious.

- 1.3 Independent claim 6 and dependent claims 7-9, if clarified as set out under Item VIII, describe an apparatus by presenting the technical features given in claims 1-4 without providing any new information concerning these technical features and without providing any new technical features. They therefore cannot be regarded

as to be inventive (Article 33(3) PCT).

2. The subject-matter of claim 10, if clarified as set out under Item VIII, does not involve an inventive step in the sense of Article 33(3) PCT, the reasons being as follows:

In accordance to D1, p. 99, left col., l. 23-32, a person skilled in the art can set up without involving any inventive step a semistructured database that manages documents, whereby each document comprises short fields for the author(s), the publication date, the title and the abstract of the document and whereby in the database there are corresponding indexes for the author(s) and the publication date and there is an index for the keywords within the titles and the abstracts of the documents. Each index is a group of index entries being identified from the documents by a common predefined criterion.

It is an obvious possibility to allow user search requests by providing an input mask, similar to masks known from electronic bookstores or from figure 1 of D1, comprising input fields for specifying the author(s), publication date and keywords of the searched documents, and where within the field for keywords the input may be given as a sequence of free selectable words and by the means shown in claim 10 or some means generally known to be equivalent thereto, keyword components are determined therefrom.

In such a system the author components, the publication date component and the determined keyword components are implicitly or explicitly allocated to the respective author slot, publication date slot and keywords slot, and the database is accessed by means as a query constructor as shown in claim 10.

Such a system comprises all features of claim 10.

Remark:

The subject-matter of independent claim 10 would seem not to be disclosed or rendered obvious in any of the documents cited in the search report if in correspondence to p. 21, l. 33 - p. 22, l. 2, p. 26, l. 8 - p. 27, l. 2, p. 27, l. 14 - p.

28, I. 34 of the description it were clear

- that the input request for information consists only of a single sequence of free selectable natural words, i.e. is not structured according to the different slots e.g. by different fields or by slot-indicating syntax words,
- that the semistructured database is having a plurality of indexes, and
- that the slot filler is arranged to allocate components to a plurality of slots whereby
 - each slot corresponds to one of the plurality of indexes and
 - for each allocated component the respective slot is determined by the slot filler out of said plurality of slots in correspondance to characteristics of said allocated component.

Problems solved by this generic input form are

- that the user is relieved from building input requests that are explicitly structured according to the different slots, and
- that there is no more need to provide and possibly distribute and update different input masks that are related to different types of search requests, which again are refering to different sets of indexes.

Regarding further documents cited in the search report:

- D3, in order to provide an easier user interface, discloses a fully structured database in which instead of inputing an formal (SQL-) query the user inputs a natural language sentence that is decomposed in components, which then are mapped onto entity attributes and relations, which not necessarily correspond to indexes. The resulting "logical form" is executed by means of relational algebra.
- D2 discloses a knowledge-based database, in which instead of inputing a command the user inputs a natural language sentence that is decomposed. The components are allocated to slots representing segments of commands, whereby all commands have the same segment structure. However, for the system no indexes are mentioned. Fitting records (commands) are located

by comparing a "goal-expression" formed by the slots with the goal-expressions corresponding to the stored command expressions.

Re Item VII

Certain defects in the international application

The features of the claims are not provided with reference signs placed in parentheses (Rule 6.2(b) PCT).

The description is not in conformity with the claims as required by Rule 5.1(a)(iii) PCT.

Contrary to the requirements of Rule 5.1(a)(ii) PCT, the relevant background art disclosed in the documents D1 and D3 is not mentioned in the description, nor are these documents identified therein.

Re Item VIII

Certain observations on the international application

Claims 1-10 do not meet the requirements of Article 6 PCT in that the matter for which protection is sought is not clearly defined, because expressions within these claims lead to contradictions or inconsistencies within the application, thereby rendering the scope of protection unclear:

- The use of the expression "record" in the claims is inconsistent to the use of the expression "record" in the description, p. 12, l. 18 - p.13, l.13. Instead, a "record" in the claims seems to have the same meaning as an "item" as used throughout the description, and is understood as such for the reasoned statement in Item V.
- Due to the unity objection of the IPEA the reference to claim 5 in claim 10 is no more acceptable. Instead, the referred features, as far as essential, should be included expressis verbis in claim 10, which is assumed for the reasoned statement in Item V.

CLAIMS

1. A method of generating an index to a semi-structured database containing a number of items, each item comprising a set of data stored in a semi-structured format, each set
5 of data including a number of related entries, the method comprising the steps of:

I) determining the presence of entries by comparing at least one set of data defining a respective item to each of a number of selection criteria, each selection criterion
10 defining one or more predetermined characteristics of a respective entry; and,

II) generating a set of indices representing a concordance between the entries determined in step (I) and the respective items.

15 2. A method according to claim 1, at least one set of data including at least a free text field having a number of entries stored as a sequence of alphanumeric characters, wherein the predetermined characteristic of at least one of the entries is the format of a number of the alphanumeric
20 characters.

3. A method according to any of the preceding claims, at least one set of data including a name entry, the predetermined characteristic of the name entry being that it is located in a name field.

25 4. A method according to any of the preceding claims, the method further comprising the step of:

III) for at least one set of data, defining any data not determined as an entry in step (I) as a free text entry.

30 5. A method according to claim 4, wherein the free text entry comprises at least one free text word defined by a sequence of alphanumeric characters, the method further comprising the steps of:

IV) determining the presence of at least one selected
35 free text word for a respective set of data by comparing the free text entry to at least one selection criterion

defining one or more predetermined characteristics of a selected free text word; and,

V) generating a set of indices representing a concordance between the selected free text words determined in step (IV) and the respective items.

6. A method according to claim 5, wherein the predetermined characteristics of the at least one selected free text word comprise a predetermined list of words.

7. A method according to any of the preceding claims, wherein the semi-structured database further comprises a number of heading fields, each heading field preceding a number of related items and including at least one heading entry, wherein the method further comprises the step of:

VI) determining the presence of heading entries by comparing each heading field to each of a number of selection criteria, each selection criteria defining one or more predetermined characteristics of a respective heading entry; and,

VII) generating a set of indices representing a concordance between the heading entries determined in step (VI) and the related items.

8. A method according to any of the preceding claims, wherein at least one index in the set of indices indicates the location of an item within the semi-structured database having a respective entry determined in any of steps (I, IV, VI).

9. A method according to any of the preceding claims, wherein at least one index in the set of indices indicates the location of each item within the semi-structured database having a respective entry determined in any of steps (I, IV, VI).

10. A method of identifying items contained in a semi-structured database having an index which has been generated according to any of the preceding claims, the method comprising the steps of:

A) generating a request for one or more items;

B) using the request and the set of indices to locate the one or more items within the semi-structured database; and

C) generating an output representative of the items.

5 11. A method according to claim 10, when dependent on claim 8 or claim 9, wherein step (B) further comprises the steps of:

B1) analysing the request and determining therefrom at least one keyterm representative of the requested items;

10 B2) comparing each keyterm with each index;

B3) selecting the indices for items which have entries including the or each keyterm; and,

B4) using the indices to determine the location of each respective item in the semi-structured database.

15 12. Apparatus for generating an index to a semi-structured database containing a number of items, each item comprising a set of data stored in a semi-structured format, each set of data including a number of related entries, the apparatus comprising:

20 a processor which determines the presence of entries by comparing at least one set of data defining a respective item to at least one of a number of selection criteria, each selection criterion defining one or more predetermined characteristics of a respective entry;

25 an index generator which generates a set of indices representing a concordance between the entries determined by the processor and the respective items; and

a data store which stores the set of indices.

30 13. Apparatus according to claim 12, each set of data including at least a free text field having a number of entries stored as a sequence of alphanumeric characters, wherein the predetermined characteristic of at least one of the entries is the format of a number of the alphanumeric characters.

35 14. Apparatus according to claim 12 or claim 13, at least one set of data including a name entry, the predetermined

characteristic of the name entry being that it is located in a name field.

15. Apparatus according to any of claims 12 to 14, wherein for at least one set of data, the processor defines any data not determined previously as an entry as a free text entry.

16. Apparatus according to claim 15, wherein the free text entry comprises at least one free text word defined by a sequence of alphanumeric characters, wherein the processor determines the presence of at least one selected free text word for a respective set of data by comparing the free text entry to at least one selection criterion defining one or more predetermined characteristics of a selected free text word; and, wherein the index generator generates a set of indices representing a concordance between the selected free text words determined by the processor and the respective items.

17. Apparatus according to claim 16, wherein the predetermined characteristics of the at least one selected free text word comprise a predetermined list of words.

18. Apparatus according to any of claims 12 to 17, wherein the semi-structured database further comprises a number of heading fields, each heading field preceding a number of related items and including at least one heading entry, wherein the processor determines the presence of heading entries by comparing each heading field to each of a number of selection criteria, each selection criteria defining one or more predetermined characteristics of a respective heading entry and wherein the index generator generates a set of indices representing a concordance between the heading entries determined by the processor and the related items, the set of indices being stored in the store.

19. Apparatus according to any of claims 12 to 18, wherein at least one index in the set of indices indicates the location of an item within the semi-structured database having a respective entry determined by the processor.

20. Apparatus according to any of claims 12 to 19, wherein at least one index in the set of indices indicates the location of each item within the semi-structured database having a respective entry determined by the processor.

5 21. Apparatus for identifying items contained in a semi-structured database having an index generated by apparatus according to any of the preceding claims, the apparatus comprising:

10 an input for receiving a request for one of more items;

a processing device which uses the request and the set of indices to locate the one or more items within the semi-structured database; and

15 an output which generates a signal representative of the items.

22. Apparatus according to claim 21, when dependent on claim 19 or claim 20, wherein the processing device is adapted to analyse the request and determining therefrom at least one keyterm representative of the requested items;

20 compare each keyterm with each index;

select the indices for items which have entries including the or each keyterm; and,

use the indices to determine the location of each respective item in the semi-structured database.

25 23. A method of accessing a database in accordance with an input request for information, the method comprising the steps of:

a) parsing the request to determine the components of the request;

30 b) determining if the request includes any verb components forming a verb or verb group;

c) if the request includes any verb components:

35 cla) determining the position of the verb or verb group within the request and, determining any subject components representing the subject of the request and any object components

representing the object of the request, using
the position of the verb or verb group;

and if the request includes no verb components:

clb) determining any components to be object
components.

d) attempting to assign at least one component to a
respective slot of a slot-and-filler request; and,

e) using the slot-and-filler request to access the
database.

24. A method according to claim 23, wherein step a)
includes the steps of:

a1) comparing each component to a predetermined hard word
list to determine hard word components, the hard word
components being assigned to the respective
grammatical class;

a2) assigning any non-hard word components to be at least
nouns;

a3) comparing each component to a predetermined verb list
to determine verb components; and,

a4) comparing each component to a predetermined
preposition list to determine prepositional
components.

25. A method according to claim 23 or claim 24, wherein
step cla) further comprises assigning any components
following the verb or verb group to be object components
and assigning any components preceding the verb or verb
group to be the subject components.

26. A method according to claim 25, when dependent on
claim 24, wherein step c) further comprises the step of:

c2) assigning any prepositional components to be modifier
components.

27. A method according to claim 26, wherein step d)
further comprises, for at least one of the slots of the
slot-and-filler request, comparing any modifier components
to a number of respective slot criteria, and assigning any
modifier component to the respective slot if it satisfies
the slot criteria.

28. A method according to any of claims 23 to 27, wherein step d) further comprises assigning any verb components to a respective slot of a slot-and-filler request.

29. A method according to claim 28, wherein the slot-and-filler request includes a transaction slot and a goods slot, the method comprising assigning any verb components to the transaction slot, and assigning any object components to the goods slot.

30. A method according to any of claims 23 to 29, wherein step d) further comprises deleting any components which are included on a predetermined list of components.

31. A method according to any of claims 23 to 30, wherein if no components can be assigned to a respective slot then the method further comprises modifying the input request and repeating steps a) to e) with the modified input request.

32. A method according to any of claims 23 to 31, wherein the at least one component is an object component.

33. A method according to any of claims 23 to 32, wherein the database is a semi-structured database having an index generated according to any of claims 1 to 9, and wherein step e) comprises the steps of:

e1) comparing each of the components assigned to a slot in the slot-and-filler request to one or more respective indices;

e2) selecting the indices for items which have entries including any of the components; and,

e3) using the indices to determine the location of each respective item in the semi-structured database.

34. Apparatus for accessing a database in accordance with an input request for information, the apparatus comprising:

an input for receiving the request;

a parser which parses the request to determine the components of the request;

a slot filler which determines if the request includes any verb components forming a verb or verb group; and, if the request includes any verb components, determines the

position of the verb or verb group within the request, and determines any subject components representing the subject of the request and any object components representing the object of the request using the position of the verb or verb group; and, if the request includes no verb components, the slot filler determines any components to be object components, and wherein the slot filler is adapted to assign at least one component to a respective slot of a slot-and-filler request; and,

10 a query constructor which uses the slot-and-filler request for accessing a database.

35. Apparatus according to claim 34, the apparatus further comprising a store for storing a predetermined hard word list, a predetermined verb list and a predetermined preposition list, the store being coupled to the parser and
15 the parser being adapted to:

 compare each component to a predetermined hard word list to determine hard word components, the hard word components being assigned to the respective grammatical
20 class;

 assigning any non-hard word components to be at least nouns;

 compare each component to a predetermined verb list to determine verb components; and,

25 compare each component to a predetermined preposition list to determine prepositional components.

36. Apparatus according to claim 35, wherein the store further stores respective slot criteria for each slot of the slot-and-filler request, and wherein the slot filler
30 compares any modifier components to the slot criteria, and assigning any modifier component to the respective slot if it satisfies the slot criteria.

37. Apparatus according to of claims 34 to 36, wherein the database is a semi-structured database having an index
35 generated by apparatus according to any of claims 12 to 21, and wherein the query constructor compares each of the components assigned to a slot in the slot-and-filler

request to one or more respective indices, selects the indices for items which have entries including any of the components and, uses the indices to determine the location of each respective item in the semi-structured database.